Interobserver Variability in Injury Severity Scoring After Combat Trauma: Different Perspectives, Different Values?

*Iain M. Smith, MB BChir, MSc, MRCSEd;
David N. Naumann, MB BChir, MRCS; Paul Guyver, MBBS, FRCS;
Jonathan Bishop, PhD; Simon Davies, BN(Hons), DipIMC RCSEd, RGN;
Jonathan B. Lundy, MD; Douglas M. Bowley, MBBS, FRCS*

ABSTRACT

Background: Anatomic measures of injury burden provide key information for studies of prehospital and in-hospital trauma care. The military version of the Abbreviated Injury Scale [AIS(M)] is used to score injuries in deployed military hospitals. Estimates of total trauma burden are derived from this. These scores are used for categorization of patients, assessment of care quality, and research studies. Scoring is normally performed retrospectively from chart review. We compared data recorded in the UK Joint Theatre Trauma Registry (JTTR) and scores calculated independently at the time of surgery by the operating surgeons to assess the concordance between surgeons and trauma nurse coordinators in assigning injury severity scores. **Methods:** Trauma casualties treated at a deployed Role 3 hospital were assigned AIS(M) scores by surgeons between 24 September 2012 and 16 October 2012. JTTR records from the same period were retrieved. The AIS(M), Injury Severity Score (ISS), and New Injury Severity Score (NISS) were compared between datasets. **Results:** Among 32 matched casualties, 214 injuries were recorded in the JTTR, whereas surgeons noted 212. Percentage agreement for number of injuries was 19%. Surgeons scored 75 injuries as "serious" or greater compared with 68 in the JTTR. Percentage agreement for the maximum AIS(M), ISS, and NISS assigned to cases was 66%, 34%, and 28%, respectively, although the distributions of scores were not statistically different (median ISS: surgeons: 20 [interquartile range (IQR), 9–28] versus JTTR: 17.5 [IQR, 9–31.5], $p = .7$; median NISS: surgeons: 27 [IQR, 12–42] versus JTTR: 25.5 [IQR, 11.5–41], $p = .7$). **Conclusion:** There are discrepancies in the recording of AIS(M) between surgeons directly involved in the care of trauma casualties and trauma nurse coordinators working by retrospective chart review. Increased accuracy might be achieved by actively collaborating in this process.

KEYWORDS: *Injury Severity Score; Abbreviated Injury Scale; trauma; surgeon; trauma nurse coordinator*

Introduction

Coding is a key process in trauma systems, providing a standardized record of individual injuries and a basis for calculation of estimates of overall injury burden. Such scores are commonly reported as demographic parameters in the trauma literature and are used to assess the performance and governance of care within and between trauma systems.[1,2] Accuracy and reproducibility in scoring, therefore, are essential.

There are four widely used scoring systems of importance in the assessment of the burden of traumatic injuries: (1) the Abbreviated Injury Scale (AIS)[3,4]; (2) a military adaptation of the AIS [AIS(M)],[5] which has been reported to predict mortality for casualties injured on military operations more accurately than the civilian version[6]; (3) the Injury Severity Score (ISS)[7]; and (4) the New Injury Severity Score (NISS).[8] These scores are calculated using objective and consistent measures. The first two systems are based on the assignment of injury severity to individual injuries within defined anatomic regions by matching the injury against detailed descriptions within the relevant dictionary; the remaining systems are derivatives of these. Illustrative injuries for the various AIS(M) grades are shown in Table 1, and Figure 1 shows

Table 1 *Examples of Injuries Assigned Various Abbreviated Injury Scale (Military Version) Scores*

Score	Severity	Example
1	Minor	Isolated rib fracture
2	Moderate	Testicular avulsion
3	Serious	Simple hemothorax
4	Severe	Below-knee traumatic amputation
5	Critical	Femoral artery injury with >20% blood loss
6	Maximum	Torso transsection

Report Documentation Page

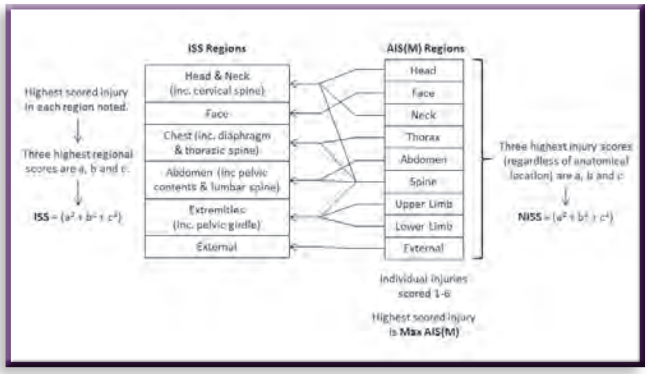
Form Approved
OMB No. 0704-0188

1. REPORT DATE **01 JUL 2015**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Interobserver Variability in Injury Severity Scoring After Combat Trauma: Different Perspectives, Different Values?**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Smith, I. M. Naumann, D. N. Guyver, P. Bishop, J. Davies, S. Lundy, J. B. Bowley, D. M.**		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **United States Army Institute of Surgical Research, JBSA Fort Sam Houston, TX 78234**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**9**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Figure 1 *Injury scoring processes.*



Note: AIS(M), Abbreviated Injury Score (Military version); ISS, Injury Severity Score; NISS, New Injury Severity Score.

the relationship between these scoring systems. NISS has been shown to be a more accurate predictor of mortality and morbidity in various clinical settings,[9–13] and it is suggested that NISS reflects military injury burden more accurately than ISS,[14] especially if derived from AIS(M).

Previous studies have demonstrated large variation in the number of injuries identified per case[15] in civilian trauma. Compared with a reference set of cases, up to 31% of injuries may be missed by trained coders.[16] Agreement for specific codes assigned to individual injuries may be as low as 39%.[16,17] This variation may lead to limited reproducibility of estimation of overall injury burden, with one study having found the probability of any two raters assigning the same ISS being as low as 28% (rising to only 51% agreement if cases are placed in bands of severity).[18] However, variation in coding of specific injuries may not compromise the reproducibility of overall measures of trauma burden: in a study of six trained raters, despite only 39% agreement for allocated AIS codes, the intraclass correlation coefficient (ICC) for ISS was "almost perfect (ICC 0.90)."[17] In contrast, a study of 10 raters found that the limits of agreement for every pairing of raters exceeded "clinically acceptable" bounds (defined by the authors as ±9 ISS or NISS units).[16] Despite these limitations, ISS remains the most commonly reported measure of injury severity in the trauma literature, and AIS(M), ISS, and NISS are routinely recorded in the JTTR. Ultimately, all three measures depend on correct identification and coding of individual injuries. The perspective of a surgeon who has seen an injury firsthand and that of a rater scoring from chart review may differ. This could lead to different scores being assigned to the same injury.

The aim of this study was to compare AIS(M) scores and derivatives assigned contemporaneously by the treating clinicians at a UK deployed military medical facility against entries for the same trauma casualties recorded in the JTTR. Based on anecdotal and personal experience of working with JTTR data, we hypothesized that scoring by physicians with clinical responsibility for patients would differ from JTTR entries resulting from chart review.

Methods

Setting

The Joint Force Medical Group Hospital at Camp Bastion was a UK-led, coalition medical facility in Helmand Province in southern Afghanistan, which provided medical support to counterinsurgency operations by the International Security Assistance Force. Military wounding patterns dominated admissions to this facility, with 75% of admissions with NISS ≥16 injured by explosive mechanisms and 23% by gunshot; 2% were injured by the blunt trauma more typical of civilian injuries.[19]

The UK JTTR

The UK JTTR is a prospectively gathered database of all patients (regardless of affiliation) who trigger activation of the trauma team at a deployed UK field hospital and of all UK Service personnel evacuated to the UK.[20] Trauma nurse coordinators (TNCs) record and code injuries after treatment at the deployed facility, based on chart review and imaging. ISS and NISS are calculated from AIS(M).

Study Design

In this preliminary, prospective, observational study, physicians coded a patient's injuries immediately after treating them. Rather than investigating possible differences in medical and nursing interpretation of written patient records, this design allowed the authors to investigate whether the definitive JTTR record accurately recorded the injuries identified by the physicians.

Selection Criteria

Patients were included in the study if they triggered trauma team activation (any casualty assigned the highest triage category or for whom team activation was initiated by any member of staff). Cases were excluded if they had not suffered injury (e.g., a patient allocated a high triage category due to a medical emergency).

Registration and Approval

This study was authorized by and registered with the Royal Centre for Defence Medicine (RCDM/Res/audit/1036/12/0264) and with the US Army Joint Combat Casualty Research Team as a performance improvement project for which institutional review board approval was not required. We adhered to the Guidelines for Reporting Reliability and Agreement Studies.[21]

Statistics

The ISS and NISS generated by each method were found to be asymmetrically distributed around their medians by using the Shapiro–Wilk test. Therefore, medians were compared by using the sign test. Reliability analysis was performed by grouping the scores according to reviewer type (clinician versus TNC) and assuming each group consisted of scores by a single reviewer. Interrater agreement was examined with the Bland–Altman limits of agreement method, with reproducibility defined as 95% of differences lying within 2 standard deviations (SDs) of the mean.[22]

Interrater agreement and reliability were further examined by using weighted κ and ICC statistics, respectively. Weighted κ statistics treat the assigned scores as ordinal data and provide an index of agreement between the two raters (clinicians and TNCs). The κ estimates were based on squared weights: the squared distance from perfect agreement determined the weight assigned to any disagreement between raters.[23]

To examine consistency rather than absolute agreement, and for comparison with similar studies,[16,17] interrater reliability of maximum AIS(M), ISS, and NISS was assessed by using ICC estimates. We assumed cases were drawn from a larger pool of casualties but that raters were fixed, and thus estimated single-measures ICC using a one-way random-effects model. ICC estimates were reported with 95% confidence intervals (CIs). Our predetermined CI width was 95%, in keeping with common practice[16] and for comparison with other published results. We used the CI to reflect the level of uncertainty associated with the estimates, not to imply statistical significance; our study was not powered to do so.

Both weighted κ and ICC were interpreted by using the arbitrary method of Landis and Koch.[24] Statistics were calculated using R, version 3.10 (R Foundation for Statistical Computing; www.r-project.org) with the MethComp[25] and irr[26] libraries.

Data Collection

During a 3-week period from 24 September 2012 to 16 October 2012, attending surgeons who operated on these patients were provided with a copy of the AIS(M) dictionary and asked to code each injury with which they had been involved. In cases where the patient underwent surgery, this coding took place as soon after the initial operation as was practical. ISS and NISS were calculated from these codes. In parallel (and independent of the study investigators), JTTR data collection continued as normal, undertaken by TNCs. JTTR scores were retrieved via a standard request for release of data analyzed after data collection was complete. The following data were identified from each source: AIS(M) for each injury, number of injuries in each body region (by AIS definition), highest AIS(M) in each body region, number of injuries for each severity grade, total number of injuries, ISS, and NISS. All matched cases were included in analysis.

Results

Data were collected by study investigators for 32 patients who received trauma-team activation during this period. Patient characteristics are shown in Table 2. The JTTR recorded "motor vehicle collision" as the mechanism of injury in three patients; however, two of these were unequivocally casualties from motor vehicles struck by improvised explosive devices (IEDs). Four additional records of "major trauma" casualties (ISS ≥16) were found in the JTTR for this period. Although these would have been expected to meet inclusion criteria, they had not been identified by the investigators. Consequently, no ratings by clinical staff had been performed, and they could not be included in analysis.

Table 2 *Patient Characteristics*

Characteristic	Data
Male sex, no. (%)	30 (94)
Age, median (IQR) [range], y	22 (21–27) [3–42]
Affiliation, no. (%)	
Afghan Security Forces	15 (47)
ISAF	11 (34)
Civilian	6 (19)
Mechanism of injury, no. (%)	
Explosion	19 (59)
Gunshot	10 (31)
Burns	2 (6)
Motor vehicle collision	1 (3)

Note: IQR, interquartile range; ISAF, International Security Assistance Force

The median ISS recorded by investigators was 20 (interquartile range [IQR], 9–28) compared with 17.5 (IQR, 9–31.5) for the JTTR (p = .7). Investigators recorded a median NISS of 27 (IQR, 12–42) while the JTTR data had a median of 25.5 (IQR, 11.5–41; p = .7). The 32 casualties had sustained 214 injuries as recorded in the JTTR, whereas the study investigators noted 212 injuries. The percentage agreement for number of injuries recorded was 19%, with a difference of up to 14 injuries per anatomic region for the entire cohort (Table 3). Injuries were recorded for 98 separate body regions in at least one of the datasets. Percentage agreement for the highest regional AIS(M) was 51%. Study investigators recorded 75 injuries scored as "serious" or greater [AIS(M) ≥3], whereas the JTTR had 68 such injuries recorded. Percentage agreement for the maximum AIS(M) assigned to cases [maxAIS(M)] was 66% (Table 4), with disagreement by one AIS(M) grade in 10 of 11 cases (91%). The remaining case differed by two grades.

Percentage agreements for identical ISS and NISS were 34% and 28%, respectively. Reproducibility criteria were

Table 3 *Total Injuries Recorded in Each Anatomic Region Compared Between Joint Theatre Trauma Registry (JTTR) and Study Investigators*

AIS Region	Injuries Recorded According to JTTR, No.	Injuries Recorded by Investigators, No.
Head	25	11
Face	23	14
Neck	0	1
Thorax	19	16
Abdomen	29	37
Spine	9	4
Upper extremity	29	35
Lower extremity	75	81
External	5	13
Other	0	0

Note: AIS, Abbreviated Injury Scale.

not met, with differences in scores lying outside 2 SDs of the mean difference in two of 32 cases (6.25%) (Figure 2 and 3). However, there was no evidence of a systematic difference in assigned ISS and NISS values within our sample.

Weighted κ indexes of interrater agreement of 0.88 ($p < .001$) "almost perfect," 0.41 ($p = .02$) "moderate," and 0.63 ($p < .001$) "substantial" were obtained for maxAIS(M), ISS, and NISS, respectively. ICC indexes of consistency in single scores were "almost perfect," with ICCs of 0.89 (95% CI, 0.78–0.94), 0.94 (95% CI, 0.88–0.97), and 0.84 (95% CI, 0.70–0.92) for maxAIS(M), ISS, and NISS, respectively. At the lower bounds of the CI, the reliability for maxAIS(M) and NISS remained "substantial" (ICC, 0.61–0.80[24]). Summary statistics for

Figure 2 *Reproducibility of Injury Severity Score (ISS) between Joint Theatre Trauma Registry (JTTR) and investigators.*

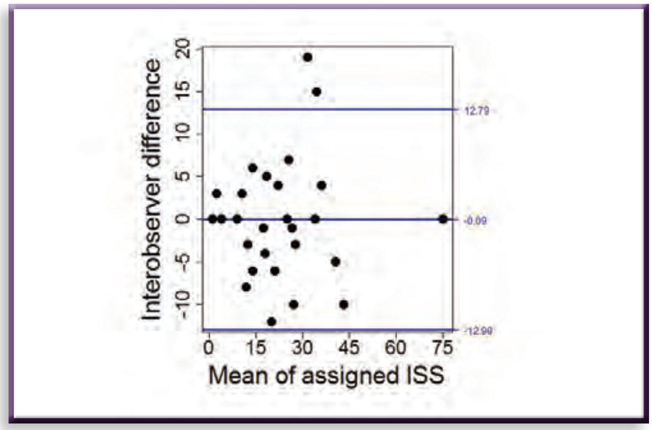
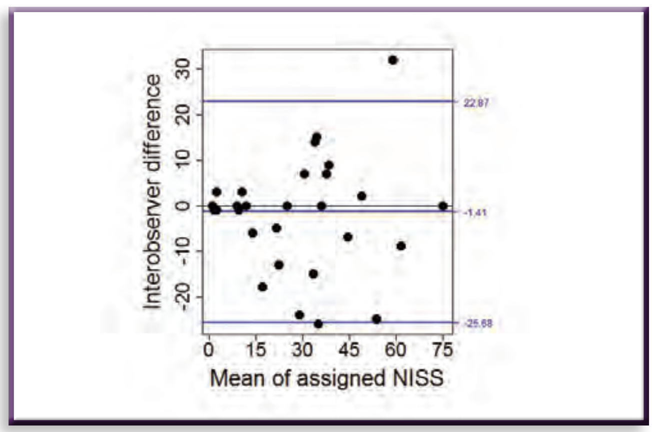


Figure 3 *Reproducibility of New Injury Severity Score (NISS) between Joint Theatre Trauma Registry (JTTR) and investigators.*



injury scores by reviewer, weighted κ, and ICC statistics are given in Table 5. The distributions of comparative ISS and NISS scores are shown in Figures 4 and 5.

Table 4 *Percentage Agreement and Interobserver Reliability of Maximum Regional Abbreviated Injury Scale (Military Version)*

Region	Agreement, %	ICC (95% CI)		Degree of Agreement at Lower CI Extreme
Head	87.5	0.908	(0.902–0.976)	Almost perfect
Face	81.3	0.746	(0.544–0.867)	Moderate
Neck	96.9	0.000	(−0.340–0.342)	None
Thorax	84.4	0.953	(0.906–0.977)	Almost perfect
Abdomen	75.0	0.881	(0.772–0.940)	Substantial
Spine	93.8	0.579	(0.297–0.769)	Fair
Upper extremity	81.3	0.883	(0.775–0.941)	Substantial
Lower extremity	68.8	0.767	(0.578–0.879)	Moderate
External	81.3	0.844	(0.707–0.921)	Substantial

Note: CI, confidence interval; ICC, intraclass correlation coefficient.

Discussion

This is the first study to assess interobserver variability of injury severity scoring in a modern combat hospital. The main finding of this study is that, during the period of observation, there were discrepancies in the AIS(M) recorded for the same patients between TNCs and treating clinicians. In addition to disagreement in injury scores, there were further differences in the total number of injuries and in both regional and total injury burdens. However, this should not be interpreted as a

Table 5 *Summary Statistics for Injury Severity Scores by Reviewer Group**

Score	Reviewer	Mean	Median	SD	Weighted κ	ICC (95% CI)
MaxAIS(M)	Surgeon	3.53	4	1.39	0.88 ($p < .001$)	0.89 (0.78–0.94)
	TNC	3.53	4	1.39		
ISS	Surgeon	22.4	17	18.5	0.63 ($p < .001$)	0.94 (0.88–0.97)
	TNC	22.5	20	18.2		
NISS	Surgeon	29.2	25.5	21.7	0.41 ($p = .020$)	0.84 (0.70–0.92)
	TNC	30.6	27.0	20.9		

Note: *Interrater agreement and reliability for each score given by weighted κ and ICC, respectively. CI, confidence interval; ICC, intraclass correlation coefficient; ISS, Injury Severity Score; MaxAIS(M), maximum Abbreviated Injury Scale (Military version); NISS, New Injury Severity Score; SD, standard deviation; TNC, trauma nurse coordinator.

Figure 4 *Comparison of Injury Severity Score (ISS) between Joint Theatre Trauma Registry (JTTR) and investigators.*

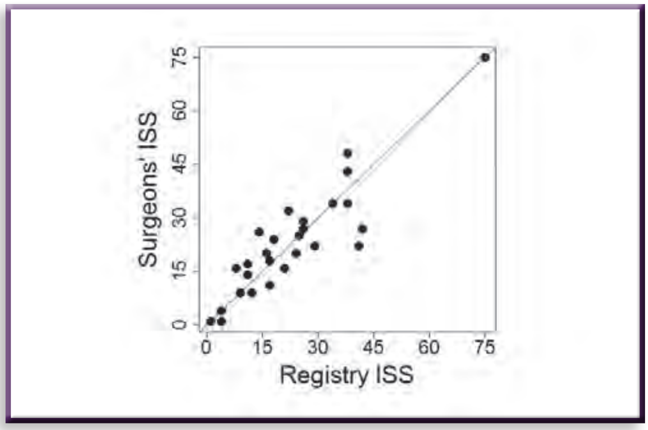
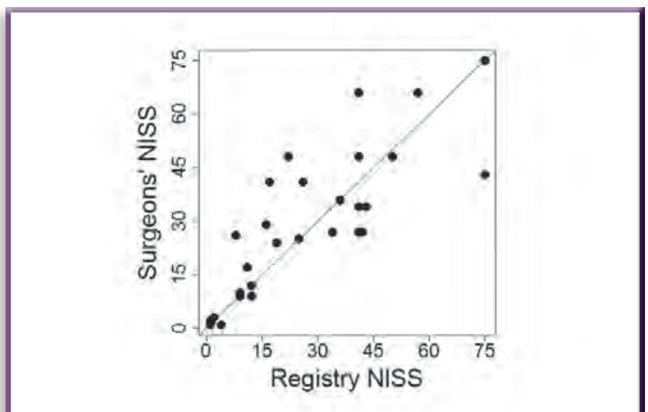


Figure 5 *Comparison of New Injury Severity Score (NISS) between Joint Theatre Trauma Registry (JTTR) and investigators.*



comment on the relative abilities of TNCs and surgeons to score injuries accurately. Rather, our results likely represent a difference in the perception of an injury visualized at the time of surgery compared with its written record in medical documentation. This may lead to differences in both the description of an injury and the interpretation of secondary features required for accurate coding (e.g., the identification of specific injuries responsible for major blood loss). No investigation of the ability of surgeons or TNCs to estimate and attribute traumatic blood loss was identified from the literature.

When comparing anatomic regions, the study investigators tended to miss injuries to the head and face. This may relate to surgeons scoring shortly after initial damage control operations, where attention had been focused on torso and limb injuries with only a provisional radiographic diagnosis of head and facial injuries available. In contrast, the definitive radiology report was available for JTTR coding. Surgeons tended to record more torso, limb, and external injuries than TNCs. This may be an effect of direct involvement with casualties (and would be consistent with previous reports that surgeon involvement improves data quality),[27] whereas conventional coding relies on the level of detail recorded in written records and the ability to match that detail to the specific descriptors within the relevant AIS dictionary. Such discrepancies between individuals when calculating injury severity scores are not unique; variation in calculation of ISS of up to 80% from the mean has been noted.[28] In a large study from Queensland, Australia, six raters independently coding 120 sets of notes achieved almost perfect agreement for ISS.[17] However, that validation exercise was based on repeated examination of standardized data rather than the comparison of visualized injuries against written records.

Despite the differences in this study between investigator and JTTR AIS(M) scores, there were no statistically significant differences between the groups in the derived ISS and NISS for these casualties. This may suggest that discrepancies in AIS(M) do not affect the predictive and prognostic function of the JTTR data. This is consistent with a previous report of “almost perfect” interrater reliability despite only 36% agreement regarding AIS.[17] However, this current study is small; statistically significant differences might be identified if a larger cohort of patients were examined as part of a study design that removed the potential for confounding discussed in the limitations section.

Our suggestion that the differences in assigned scores arise from different exposures to the casualty is consistent with previous findings that involving trauma surgeons in coding at a Level I trauma center resulted in coding amendments in 5.2% of cases scored by nonsurgeon coders alone, with 71% of these amendments being an upward revision of ISS.[27] If accurate scoring in the military context is to be achieved, then active involvement of the surgeon will be essential. The authors consider it vital that deployed TNCs feel able to seek clarification from the operating surgeon, while surgeons could contribute to improvement by describing injuries in language consistent with the dictionary when writing the operative record. Together, the TNC and the surgeon should generate a consensus injury description to describe the injury in terms of the most accurate "fit" in the scoring dictionary. Existing collective predeployment training provides an ideal opportunity to establish collaboration by the inclusion of suitable casualty scenarios.[29]

A further potential contributor to our findings is that UK TNCs are trained for specific operations but rarely have significant exposure to trauma data capture before deployment and are unlikely to return subsequently to that role. Detailed training in anatomy, radiology report writing, and injury description is required to undertake the TNC role effectively. Establishing a Defence Medical Services TNC cadre, working within the civilian UK Major Trauma Networks, would allow participation in the ongoing UK Trauma Audit & Research Network (TARN) activity, leading to increased expertise in the matching AIS(M) descriptions to computed tomography scan reports, surgical operation notes, and postmortem studies. While civilian and military wounding patterns differ, such exposure would enable TNCs to deploy with greater experience in trauma audit and governance.

A secondary finding of this study is that the JTTR may not accurately record mechanism of injury data when an IED strike to a military vehicle precipitates a rollover or other accident in which patients sustain injuries typical of a more conventional motor vehicle incident. Finally, the failure of investigators to identify and record all eligible patients almost certainly relates to their primary responsibility of providing surgical care to the wounded. This reinforces the importance of the deployed hospital establishment including registry personnel whose role is dedicated to information capture. This principle should also apply to data capture for research projects in the deployed environment.

Study Limitations

Data collection for this study was conducted by two British military surgical residents in the second half of a 7-week deployment. This brief operational tour limited the sample size that could be collected, but it was considered sufficient for a preliminary investigation. It is possible that this period was unrepresentative of normal activity and that greater casualty flow might have affected the accuracy of physician coding by reducing the time available to match injuries to detailed descriptions in the AIS(M) dictionary. However, the experience of the senior authors from multiple, longer deployments suggests that the activity level was not unusual.

A further limitation is that our selection criteria focused on patients likely to have more severe injuries. This could lead to greater apparent disparity between physicians and registry staff, as external injuries are predominantly minor and would be easily recognized as such by both groups. However, these more complex injuries are precisely those in which detail may be lost in the written record and, thus, are most relevant to our hypothesis.

This study did not verify the accuracy of the surgeons' assessment of injuries, and none of our investigators had been formally certified in AIS methodology. However, the UK surgical training syllabus requires competence in the use of trauma scoring systems,[30] and previous work has found no significant difference in performance between trained and untrained raters.[18]

This study does not identify whether the difference in assigned scores arises from variability between individual raters, from documentation that does not convey a sufficient level of detail to permit accurate retrospective coding, or from a different interpretation of injuries arising from the different professional backgrounds of physicians and registry personnel. Because our study was intended to compare the perceptions of the treating clinicians with registry output, it was not possible to have the same clinicians score each case. Logistical constraints meant the sample size was insufficient to perform multilevel model-based analysis of cross-classified data[31] (not all cases assessed by all reviewers) that could account for this. This should be studied further to refine registry process and maximize data quality and analysis.

Conclusion

This study highlights discrepancies in injury severity scoring between surgeons and TNCs, with surgeons recording more torso, limb, and external injuries, and TNCs including more injuries to the head and face. Because injury scoring is vital to the analysis of the efficacy of interventions at all stages of the trauma care pathway, from point of wounding to discharge from rehabilitation, it is important to minimize such differences. Both TNCs and treating clinicians should be actively involved in the injury scoring of patients. Collaboration will improve both accuracy and reliability.

Acknowledgments

The authors thank the various consultant surgeons and radiologists who participated in this study. The Clinical Information Exploitation Team and Defence Statistics

Health are thanked for collecting, collating, and identifying the JTTR data used in this report.

Disclaimer

This report presents independent research supported in part by the National Institute for Health Research (NIHR) Surgical Reconstruction and Microbiology Research Centre (a partnership among University Hospitals Birmingham National Health Service [NHS] Foundation Trust, the University of Birmingham, and the Royal Centre for Defence Medicine). The views expressed are those of the author(s) and not necessarily those of the UK Ministry of Defence, US Department of Defense, the NHS, the NIHR, or the UK Department of Health.

Disclosure

The authors have no financial relationships relevant to this article to disclose.

References

1. **Russell RJ, Hodgetts TJ, McLeod J, et al.** The role of trauma scoring in developing trauma clinical governance in the Defence Medical Services. *Philos Trans R Soc Lond B Biol Sci.* 2011;366:171–191.
2. **Stevenson M, Segui-Gomez M, Lescohier I, et al.** An overview of the injury severity score and the new injury severity score. *Inj Prev.* 2001;7:10–13.
3. **Gennarelli TA, Wodzin E.** AIS 2005: a contemporary injury scale. *Injury.* 2006;37:1083–1091.
4. **Gennarelli TA, Wodzin E, eds.** *Abbreviated Injury Scale 2005: update 2008.* Barrington, IL: Association for the Advancement of Automotive Medicine; 2008.
5. **Champion HR, ed.** *Military Abbreviated Injury Scale 2005.* Silver Springs, MD: SimQuest LLC; 2007.
6. **Champion HR, Holcomb JB, Lawnick MM, et al.** Improved characterization of combat injury. *J Trauma.* 2010;68:1139–1150.
7. **Baker SP, O'Neill B, Haddon W Jr, et al.** The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma.* 1974;14:187–196.
8. **Osler T, Baker SP, Long W.** A modification of the injury severity score that both improves accuracy and simplifies scoring. *J Trauma.* 1997;43:922–925.
9. **Balogh Z, Offner PJ, Moore EE, et al.** NISS predicts postinjury multiple organ failure better than the ISS. *J Trauma.* 2000;48:624–627.
10. **Balogh ZJ, Varga E, Tomka J, et al.** The new injury severity score is a better predictor of extended hospitalization and intensive care unit admission than the injury severity score in patients with multiple orthopaedic injuries. *J Orthop Trauma.* 2003;17:508–512.
11. **Brenneman FD, Boulanger BR, McLellan BA, et al.** Measuring injury severity: time for a change. *J Trauma.* 1998;44: 580–582.
12. **Harwood PJ, Giannoudis PV, Probst C, et al; Polytrauma Study Group of the German Trauma Society.** Which AIS based scoring system is the best predictor of outcome in orthopaedic blunt trauma patients? *J Trauma.* 2006;60:334–340.
13. **Sullivan T, Haider A, DiRusso SM, et al.** Prediction of mortality in pediatric trauma patients: new injury severity score outperforms injury severity score in the severely injured. *J Trauma.* 2003;55:1083–1087.
14. **Smith J, Hodgetts T, Mahoney P, et al.** Trauma governance in the UK defence medical services. *J R Army Med Corps.* 2007;153:239–242.
15. **MacKenzie EJ, Shapiro S, Eastham JN.** The Abbreviated Injury Scale and Injury Severity Score. Levels of inter- and intrarater reliability. *Med Care.* 1985;23:823–835.
16. **Ringdal KG, Skaga NO, Hestnes M, et al.** Abbreviated Injury Scale: not a reliable basis for summation of injury severity in trauma facilities? *Injury.* 2013;44: 691–699.
17. **Neale R, Rokkas P, McClure RJ.** Interrater reliability of injury coding in the Queensland Trauma Registry. *Emerg Med (Fremantle).* 2003;15:38–41.
18. **Zoltie N, de Dombal FT.** The hit and miss of ISS and TRISS. Yorkshire Trauma Audit Group. *BMJ.* 1993;307:906–909.
19. **Smith IM.** Coagulopathy after military trauma: the impact of a pre-hospital major haemorrhage protocol [dissertation]. Swansea, Wales: Swansea University; 2012.
20. **Brodie S, Hodgetts TJ, Ollerton J, et al.** Tourniquet use in combat trauma: UK military experience. *J R Army Med Corps.* 2007;153:310–313.
21. **Kottner J, Audigé L, Brorson S, et al.** Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64:96–106.
22. **Bland JM, Altman DG.** Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;327:307–310.
23. **Agresti A.** *Categorical data analysis.* 2nd ed. New York, NY: John Wiley & Sons, Inc.; 2002.
24. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
25. **Carstensen B, Gurrin L, Ekstrom C, et al.** *MethComp: functions for analysis of agreement in method comparison studies.* R package version 1.22. 2013. http://CRAN.R-project.org/package=MethComp. Accessed 10 June 2014.
26. **Gamer M, Lemon J, Fellows I, et al.** *irr: various coefficients of interrater reliability and agreement.* R package version 0.84. 2012. http://CRAN.R-project.org/package=irr. Accessed 29 October 2014.
27. **Mikhail JN, Harris YD, Sorensen VJ.** Injury severity scoring: influence of trauma surgeon involvement on accuracy. *JTrauma Nurs.* 2003;10:43–47.
28. **Waydhas C, Nast-Kolb D, Trupka A, et al.** [Trauma scores: reproducibility and reliability]. *Unfallchirurg.* 1992;95:67–70.
29. **Davies TJ, Nadin MN, McArthur DJ, et al.** Hospex 2008. *J R Army Med Corps.* 2008;154:195–201.
30. **Joint Committee on Surgical Training.** *The intercollegiate surgical curriculum.* October 2013. www.iscp.ac.uk/documents/syllabus_GS_2013.pdf. Accessed 16 November 2014.
31. **Goldstein H.** *Multilevel statistical models.* 3rd ed. London, UK: Arnold; 2003.

Maj Smith is a general surgical resident in the UK Army Reserve, with deployed surgical experience in both Iraq and Afghanistan. He was previously a prehospital squadron second-in-command. He is currently a Clinical Research Fellow at the Royal Centre for Defence Medicine and the NIHR Surgical Reconstruction & Microbiology Research Centre, Queen Elizabeth Hospital; and is affiliated with the 205 (Scottish) Field Hospital in Glasgow, UK. He may be contacted c/o NIHR SRMRC, Office 4.36, East Block, 4th Floor, Research & Development—University Hospitals Birmingham NHS Foundation Trust, Queen Elizabeth Hospital, Queen Elizabeth Medical Centre, Birmingham, B15 2TH, UK. E-mail: iainmsmith@doctors.net.uk.

Maj Naumann is a resident in general surgery with an interest in trauma. He deployed in a prehospital role in Afghanistan, where he provided combat medical cover for frontline patrol bases and checkpoints in Helmand Province. He is affiliated with the Royal Centre for Defence Medicine, Queen Elizabeth Hospital, Birmingham, UK.

Surg Cdr Guyver is a commando-trained orthopaedic surgeon. He has deployed afloat as the lead Medical Officer for STANAVFORMED with UK Commando forces in both Iraq and Afghanistan and as orthopaedic resident at Camp Bastion. He was recently appointed Consultant Orthopaedic and Trauma Surgeon. He is affiliated with the Institute of Naval Medicine, Hampshire, UK.

Dr Bishop is a civilian medical statistician at the NIHR Surgical Reconstruction & Microbiology Research Centre, Queen Elizabeth Hospital, Birmingham, UK. He has provided statistical expertise to a number of military research projects.

Lt Col Davies is SO1 BATLS in the UK Army Reserve. He has deployed on five operational tours in both Iraq and Afghanistan, several times in the role of Trauma Nurse Coordinator. He is currently the Major Trauma Coordinator at the Royal Stoke University Hospital in Stoke on Trent, University Hospitals of North Midlands NHS Trust, Stoke on Trent, UK.

LTC Lundy is a trauma/burn/general surgeon at the US Army Institute of Surgical Research, Fort Sam Houston, Texas.

Lt Col Bowley is a consultant general and colorectal surgeon who has served with UK Airborne Forces. He is Consultant Advisor in Surgery to the Director General Army Medical Services. He is affiliated with the Royal Centre for Defence Medicine, Queen Elizabeth Hospital, Birmingham, UK.